



PREDICCIÓN DEL FRAUDE CON TARJETAS PARA UNA ENTIDAD FINANCIERA
A TRAVÉS DEL MODELO ARIMAX

Molano Díaz Yuly Alexandra
Correa Ramos Yamir

Fundación Universitaria Los Libertadores
Departamento de Ciencias Básicas
Especialización en estadística aplicada

Bogotá D.C.
2016



PREDICCIÓN DEL FRAUDE CON TARJETAS PARA UNA ENTIDAD
FINANCIERA A TRAVÉS DEL MODELO ARIMAX

Molano Díaz Yuly Alexandra
Correa Ramos Yamir

Asesor estadístico: Juan Camilo Santana
Asesor estadístico: Diana Walteros

Fundación Universitaria Los Libertadores
Departamento de Ciencias Básicas
Especialización en estadística aplicada

Bogotá D.C.
2016

Nota de Aceptación

Firma del presidente del jurado

Firma del Jurado

Firma del Jurado

Bogotá, D.C 31 Julio del 2016

Las Directivas de la Universidad de
Los Libertadores, los jurados calificadores y el cuerpo
Docente no son responsables por los
criterios e ideas expuestas En el presente documento.
Estos corresponden únicamente a los autores

Tabla de contenido

Resumen	8
Abstract	8
Capítulo 1 Introducción.....	9
Objetivo	10
Objetivos Específicos	10
Justificación.....	10
Capítulo 2 Marco de Referencia.....	11
Capítulo 3 Marco Teórico	16
Capítulo 4 Marco metodológico.....	19
Capítulo 5 Análisis y Resultados.....	22
Conclusiones	37
Referencias	38

Lista de Tablas

Tabla 1. Descripción de las modalidades, 14.

Tabla 2 Análisis Descriptivo, 23.

Tabla 3. Descripción de los posibles Modelos, 26.

Tabla 4. Predicción del fraude para el 2016 modelo ARIMA, 27.

Tabla 5. Ajuste Datos Atipicos Outliers ("AO", "LS", "TC"), 29.

Tabla 6. Estimación Modelo ARIMAX, 30.

Tabla 7. Pruebas realizadas modelo ARIMAX, 32.

Tabla 8. Proyección del fraude para el año 2016 en la entidad financiera, 32.

Tabla 9. Proyección de fraude v/s cifras de la entidad, 33.

Tabla 10. Promedio de fraude por mes y su proyección, 36

Lista de Figuras

- Figura 1. Comportamiento del fraude según su modalidad, 14.
- Figura 2: Clientes afectados por Fraude, 19.
- Figura 3: Comportamiento del Fraude mensual entre el periodo 2011 – 2015, 22.
- Figura 4: histograma con función de densidad de la serie de Fraude del Entidad financiera, 23.
- Figura 5: Función de Autocorrelación, 25.
- Figura 6: Función de Autocorrelación Parcial, 25.
- Figura 7: Análisis de residuales del modelo, 27.
- Figura 8: Ajustado Observado y Predicción, 28.
- Figura Serie 9: inicial y con datos ajustados, 30.
- Figura 10: Resultado grafico del Modelo ARIMAX, 31.
- Figura 11: Resultado grafico de la Predicción del fraude para el 2016, 33.
- Figura 12: Predicción del fraude versus comportamiento mensual de la serie 2011-2016, 35

Lista de Anexos

Anexo 1 Base del fraude total.

PREDICCIÓN DEL FRAUDE CON TARJETAS PARA UNA ENTIDAD FINANCIERA A TRAVÉS DEL MODELO ARIMAX

**** Yuly Molano Diaz, Yamir Correa Ramos.**

RESUMEN

Este documento tiene como objetivo analizar la serie que refleja el comportamiento del fraude mensual con tarjetas para una Entidad Financiera, así como también realizar un modelo ARIMAX(p,d,q) e identificar cual podría ser el fraude para el año 2016, en busca de Analizar el mismo para evaluar mejores prácticas, demostrando la capacidad de reacción del banco y su posición frente a las otras entidades del sector.

Los resultados muestran que el mejor modelo para predecir el comportamiento del fraude en la entidad es un ARIMAX (0,1,1) mediante la metodología de series de tiempo, aun cuando existen factores de volatilidad en la serie.

Palabras claves: ARIMAX, Entidad Financiera, Fraude.

CARD FRAUD PREDICTION FOR A FINANCIAL ENTITY THROUGH THE ARIMAX MODEL

**** Yuly Molano Diaz, Yamir Correa Ramos.**

ABSTRACT

This document aims to analyze the series that reflects the behavior monthly of card fraud for a financial entity, as well as making an ARIMAX model (p, d, q) and identify which could be fraud in 2016, seeking analyze it to evaluate best practices, demonstrating the capacity to react of the bank and its position against other entities in the sector.

The results show that the best model to predict the behavior of fraud in the entity is an ARIMAX (0,1,1) , although there are factors volatility in the series.

Keywords: ARIMAX, Financial Entity, Fraud.

Capítulo 1. Introducción

En la actualidad el fraude es una problemática que preocupa tanto a los Bancos como a los clientes del mismo, en primera medida es importante conocer que son varios los factores que han aumentado el riesgo (probabilidad de ocurrencia de un evento negativo) de fraude bancario; entre estos están, el uso de nuevas tecnologías en el sector bancario por parte de los clientes para realizar diversas transacciones las cuales en gran medida han permitido a los delincuentes acceder con mayor facilidad a la información bancaria del mismo, facilitando el fraude ante los clientes; el aumento en el uso de internet en la población; Sin embargo, este crecimiento ha traído un crecimiento también en el riesgo de fraude, en general ha facilitado la propagación de ataques cibernéticos mediante virus; múltiples estrategias por el defraudador tales como la falsificación de la banda magnética, el cambiazó, la suplantación, el robo y el extravío; la debilidad en los procesos de aseguramiento de la información de las tarjetas por parte de la entidad financiera, de los clientes y comercios, así como también se manejan más políticas reactivas que proactivas.

Teniendo en cuenta lo anterior, se han generado diversas estrategias para mitigar el fraude, entre estas la utilización de modelos estadísticos que permitan predecir determinados factores del mismo, dichos modelos están basados principalmente en sistemas analíticos como la minería de datos, que permiten el procesamiento de información masiva y la extracción de conocimiento de la misma, por ejemplo se analiza información de transacciones realizadas por los clientes versus las transacciones fraudulentas, con el fin de encontrar patrones diferenciadores que permitan detectar el comportamiento del defraudador. Aparentemente este tipo de análisis es el que predomina en el sector, dejando a un lado otro tipo de técnicas que podrían aportar información valiosa y diferente, en temas relacionados con fraude, como por ejemplo el análisis de series de tiempo.

Sin embargo a medida que avanza la facilidad para realizar transacciones bancarias, aumenta el riesgo y de mantener los mismos controles, sin agregar mayor tecnología y análisis estadístico del fraude este tiende a crecer, como se mencionó anteriormente el fraude se ha desplazado al fraude por Internet, esto debido a que la información solicitada para realizar transacciones por Internet es mínima, es más rápido y eficiente realizar transacciones en menor tiempo y sin ningún tipo de riesgo de ser encontrados realizando la transacción, el fraude se incrementaría, afectando la imagen corporativa de la entidad. Una de las posibles soluciones o que permiten reducir el fraude es invertir en tecnología, esto con el fin de ser más rápidos en la forma de reaccionar ante los ataques de fraude por Internet, sin descuidar la educación financiera del manejo y custodia del plástico para los clientes, así como también se encuentra ligado el análisis de la información el cual incluye el desarrollo de modelos estadísticos que permitan

pronosticar los posibles clientes afectados y el valor futuro posible; estos análisis son de gran importancia para las entidades financieras ya que permiten visualizar la información agrupada que contiene el 70% del fraude, como cuánto y porque el cliente es afectado, con el fin de establecer patrones de comportamiento de los delincuentes he identificar las variables que influyen en el fraude por cada modalidad. Teniendo en cuenta lo anterior el presente estudio formula el diseño de un modelo de predicción del fraude a partir del comportamiento histórico de los datos de la entidad financiera, dada la necesidad de conocer la conducta del fraude (tanto en montos como en variables que ayudan a la predicción del mismo) en los clientes de la entidad.

Objetivo

Diseñar un modelo que permita predecir el comportamiento del fraude para la entidad financiera.

Objetivos Específicos

- a. Identificar los principales patrones de comportamiento del fraude con tarjetas en la entidad financiera.
- b. Predecir el comportamiento del fraude por un periodo de un (1) año para la entidad financiera en análisis.

Justificación

Las series de tiempo ofrecen fortalezas para analizar el comportamiento del fraude debido a que son más sencillos de utilizar, permiten observar la variable a lo largo del tiempo desempeñando un papel importante en el análisis requerido para el pronóstico de eventos futuros, utilizar este análisis admite considerar además, que los factores que han venido ocurriendo en la serie o período a evaluar seguirán influenciado del mismo modo en nuestro escenario futuro, brindando una visión de la realidad analizada y permitiendo ver cualquier cambio fuerte o inesperado en algún periodo.

Por otra parte el comportamiento de las series de tiempo tiene en cuenta la tendencia, la variación cíclica y la variación estacional, dicho análisis nos permite garantizar la validez de nuestros resultados, obtener conclusiones mucho más precisas y exactas que nos permite predecir el comportamiento del fenómeno del fraude con un nivel de confiabilidad y precisión muy elevado, este estudio es visto como una estrategia que permite tener una visión futura del comportamiento del fraude según sus modalidades para el producto tarjeta de crédito de la entidad financiera, dado que las tarjetas de crédito actúan como uno de los medios de pago más importantes en la economía de Colombia y permiten acceder a bienes y servicios sin necesidad de contar con el dinero en el momento a miles de usuarios del sistema financiero.

Capítulo 2. Marco de Referencia

El sector financiero en Colombia representa un factor fundamental en la economía del país, de allí que el gobierno invite a los ciudadanos a la utilización de los servicios financieros, “En muchas de las discusiones que han rodeado estas agendas incluso se afirma que en el contexto actual la exclusión financiera equivale a una forma de exclusión social” (Subirats, Gomà y Brugué, 2005, p.2). Según cifras de la Superintendencia Financiera de Colombia a enero del 2016 los activos del sistema financiero alcanzaron un valor de \$1,264.29 billones y las utilidades “de las instituciones de crédito ascendieron a \$639.59mm, de las cuales \$602.78mm correspondieron a los bancos” (Superintendencia Financiera, 2016, p.4). La expansión que ha tenido el comercio se ha visto favorecido por herramientas que permiten realizar transacciones sin necesidad de acudir a las instituciones bancarias, un ejemplo son los cajeros electrónicos, la banca móvil o los puntos de venta mecanizados e integrados a la red del sistema financiero; cada vez es más común ver un dinamismo en estos procesos por parte de las entidades, donde la banca electrónica virtual juega un papel fundamental en el objetivo de facilitar el consumo de bienes y servicios, propiciando que los consumidores interactúen de forma más ágil accediendo en cualquier momento y lugar a sus finanzas y por ende a sus productos, muchos autores afirma que el cambio tecnológico sucede a la par de los cambios en la economía.

En los últimos años los cambios son notorios en el sistema financiero y en la forma de ofrecer los productos y servicios a los distintos clientes, aquí ya no solo se tiene en cuenta el producto como tal, se tiene en cuenta la imagen de la entidad, el servicio, la rapidez, la presentación personal de la personas que prestan el servicio e incluso el aspecto geográfico, aspectos que hacen que las entidades o instituciones compitan por mejorar la eficiencia buscando reducir costos que favorezcan los márgenes de ganancia, que permitan ser más eficientes frente a los deberes con los usuarios, en establecer políticas y reglas necesarias para mantener la confianza de los clientes con el sistema financiero, por lo tanto cada entidad está en el deber de disponer de la información con calidad, disponibilidad inmediata en los precios de los productos y servicios ofrecidos.

Ser más ágiles en prestar los servicios puede a su vez favorece en algunos aspectos a la delincuencia, por ello hoy Superintendencia Financiera de Colombia, a la cual se le encargó la función de velar por la estabilidad general del sistema financiero. “La composición del sistema financiero y su regulación se encuentran comprendidas en el Estatuto Orgánico del Sistema Financiero –EOSF– y las normas que regulan el mercado de valores y la regulación que fusionó la Superintendencia Bancaria y la Superintendencia de Valores actualmente consolidadas en el Decreto 2555 de 2010”(Dovale y Roa, 2012, p.12).

En el país las entidades financieras como los bancos ofertan productos de captación como Cuentas de Ahorro, Cuentas Corrientes y certificados de depósito a término (CDT'S o CDAT'S) administrando de esta manera recursos de los clientes y en contraprestación pagando una tasa de interés acordada.

Las cuentas de ahorro varían de acuerdo a las políticas de la entidad financiera y son ofrecidas con el propósito de obtener recursos, las cuentas corrientes tienen la particularidad de ser un producto que no genera intereses, y los CDT'S respaldan un depósito en un plazo acordado entre el banco y el cliente generando un rendimiento de acuerdo a la tasa de interés pactada.

Por otra parte los Bancos también prestan servicios financieros y productos de consumo a través de productos regidos por la Superintendencia Financiera de Colombia como, Tarjetas de Crédito, Créditos de Consumo, Créditos Educativos, Créditos de Vehículo, Créditos Hipotecarios destinado a la compra de bienes inmuebles, Créditos de libre inversión destinados para la adquisición de cualquier tipo de bien y créditos dirigidos a las empresas ya sea para mejorar procesos o para la operación misma, según cifras de la superintendencia financiera de Colombia para enero del 2016

La tasa de interés promedio para las tarjetas de crédito fue 28.34%, 118pb por debajo de la tasa de usura vigente durante el mes (29.52%). Los otros productos de consumo registraron una tasa promedio de 18.63%, superior en 90pb a la observada en diciembre de 2015 bancos” (2016, p.14).

El interés bancario es certificado por la Superintendencia Financiera de Colombia, la cual requiere a las entidades del sistema financiero una tasa promedio de acuerdo a la modalidad de crédito correspondiente, que de acuerdo con el artículo 11.2.5.1.2 del Decreto 2555 de 2010, la Ley 45 de 1990. “Por la cual se expiden normas en materia de intermediación financiera, se regula la actividad aseguradora, se conceden unas facultades y se dictan otras disposiciones” (Dovale y Roa, 2012, p.12).

Es importante mencionar que los medios de pago más utilizados por los clientes es el efectivo, cheque o con tarjetas débito y crédito, en los pagos con tarjetas de crédito existen cuatro franquicias que funcionan como intermediarias MasterCard, Visa, American Express y Dinner, las tarjetas de crédito en Colombia vienen desarrollándose cada vez más de acuerdo con la facilidad de acceso al producto, con el buen entorno económico presentado y las condiciones de mercado, lo que lleva a que en la actualidad sea uno de los medios de pago de mayor utilización en nuestra economía, así mismo se ha vuelto cada vez más fuerte la competencia debido a que es un producto muy rentable, además genere mayores volúmenes, mayores transacciones y por consecuencia, mayores ganancias, por lo anterior también implica mayores riesgos en cuanto a la utilización del mismo, las entidades oferentes del producto hacen esfuerzos grandes no solamente en poner al alcance la tarjeta de crédito al cliente, sino en tener en cuenta aspectos como beneficios, tarifas, programas de fidelización, características del plástico y tecnología; aspectos que prácticamente definen el desempeño de la colocación y uso de tarjetas de crédito en el país, es así que:

De acuerdo con el Banco de la República el uso de tarjetas débito y crédito ha venido creciendo de manera continua a lo largo de la década. En 2000 el número promedio anual de tarjetas de crédito fue 1.951.646 y en 2013 el total de tarjetas al cierre del año correspondía a 13.475.458, lo que implica un crecimiento durante los 13 años de 704%” (Sánchez, 2014, p. 7).

Lo anterior y teniendo de la mano a la tecnología la cual se ha vuelto la principal herramienta de las empresas que abre la posibilidad de la utilización de los servicios y permite la utilización de canales como el Internet, la banca móvil, la seguridad en tarjetas con el Chip, que vienen en auge en el sector Bancario, siempre en aras de la estabilidad del sistema y la seguridad de los productos, en este punto la Superintendencia Financiera de Colombia con la circulas 052 de 2007, la más reciente “la cual incrementa los estándares de seguridad y calidad para el manejo de la información a través de medios y canales de distribución de productos y servicios que ofrecen a sus clientes y usuarios las entidades vigiladas por esta Entidad.”(Superintendencia Financiera de Colombia, 2007, p. 1). Está circular fija para los siguientes medios:

Tarjetas débito y crédito y los Canales Oficinas, Cajeros Automáticos (ATM), Receptores de cheques, Receptores de dinero en efectivo, POS (incluye PIN Pad), Sistemas de Audio, Respuesta (IVR), Centro de atención telefónica (Call Center, Contact Center), Sistemas de acceso remoto para clientes (RAS), Internet y Dispositivos móviles (Superintendencia Financiera de Colombia, 2007, p. 1)

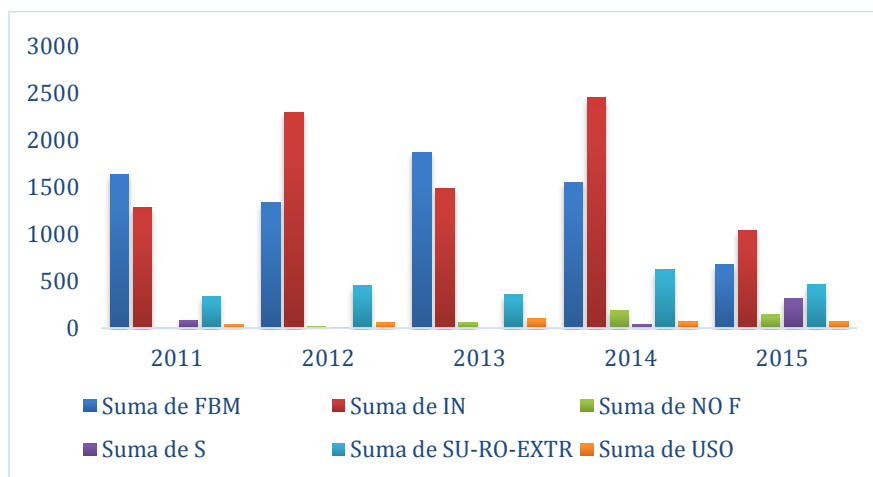
Políticas y normas encaminadas a prevenir, evitar y proteger tanto la información financiera como los recursos de los clientes, que también buscan mayor interacción con el cliente en cuanto se le informe en línea la utilización de su producto, la seguridad del mismo y brindar la información para tomar medidas preventivas con el fin de evitar que terceros puedan tener acceso a la información; debido a esto cobra especial relevancia para las entidades financieras estudiar la evolución del fraude con tarjetas crédito y débito, de las franquicias MasterCard y Visa, en este caso debido a que la entidad financiera en estudio maneja solo estas dos franquicias, durante el período comprendido entre 2011 y 2016, para comenzar es necesario tener en cuenta que la industria financiera en Colombia para el año 2.010 inicia un nuevo ciclo con sus tarjetas de crédito y débito, a partir del cambio de tecnología migrando al Chip, de acuerdo con la norma de la Superintendencia financiera a través de la circular 042 del 2.012, aquellas tarjetas que no cuenten con la tecnología y presenten fraude bajo la modalidad de Falsificación de Banda Magnética (FBM), este será asumido por la entidad que no cuenta con dicha tecnología, es así como la industria financiera migro a esta tecnología, el efecto inmediato de la implementación del Chip en el país y las nuevas dinámicas del mercado frente a la facilidad, uso y manejo de las formas de pago, fue el desplazamiento del fraude a otros países que no contaban con dicha tecnología o la migración del fraude con tarjetas de crédito mediante el canal Internet y para el producto tarjetas débito el robo de las mismas o utilización en el exterior. También es importante mencionar que los comercios fueron

implementando nuevas opciones y facilidades para que los tarjetahabientes puedan hacer uso del canal Internet para las tarjetas de crédito, entre ellos el viernes negro, ciber-lunes etc; lo que hace que las personas que utilizan el producto utilicen dicho servicio, para la adecuada comprensión del documento es necesario identificar la nomenclatura utilizada por la entidad financiera en estudio, para cada una de sus modalidades de fraude:

Tabla 1. Descripción de las Modalidades

SIGLA	DESCRIPCIÓN
FMB	Falsificación de Banda Magnética.
IN	Fraude por Internet.
S	Fraude por Suplantación de persona.
NO F	Casos que no corresponden a fraude, sin embargo, fueron pagados a los tarjetahabientes y reportados ante las franquicias.
SU-RO-EXTR	Fraude por sustitución de la tarjeta, robo o extravío de la misma.
USO INDEBIDO	Fraude correspondiente a la inapropiada utilización de la tarjeta por parte del cliente.

La nomenclatura de la tabla 1 facilitara observar en el figura 1 el comportamiento del fraude a través de los años, con el fin de ver la magnitud de cada una en el fraude total reportado por la entidad.



Fuente: Entidad Financiera, cifras de fraudes expresados en Millones.

Figura 1: Comportamiento del Fraude según su modalidad.

En la figura 1 se puede observar el incremento del fraude a través de los años por el canal Internet, siendo el año 2014 el más afectado, en la figura también se observa que las modalidades sustitución de tarjeta robo y extravío se mantienen a través de los años sin presentar grandes cambios, estos resultados nos permiten inferir que como los defraudadores no pueden falsificar la banda magnética o más bien

utilizarla en Colombia entonces buscan quitarle las tarjetas a los tarjetahabientes, es de señalar que esta modalidad está comprendida en un 90% en el producto tarjeta débito y para llevar a cabo el fraude es necesario contar con el PIN de la misma. De acuerdo con la implementación del Chip se esperaba una disminución más significativa del fraude bajo la modalidad FBM, sin embargo, los resultados solo muestran esta disminución para el año 2015, dicho resultado obedece a la **NO** implementación del Chip en el país Estados Unidos, el cual es el principal país utilizado por los delincuentes para llevar a cabo el fraude bajo la modalidad de FBM, para el periodo comprendido entre el 2011-2014.

En la gráfica también se puede observar que la modalidad no fraude, ha ido creciendo con el pasar de los años, esto es normal debido a que la Entidad ha ido implementando mejores prácticas para identificar reclamos injustificados, algunos de buena fe y otros por mala fe del tarjetahabiente.

Por otra parte, se observa que para la modalidad suplantación de persona no se presentaron reclamaciones durante el periodo 2012-2013 y vuelve a coger fuerza en el año 2015 con una participación del 11%.

La NO migración de chip por parte de Estados Unidos durante el periodo de análisis, es un factor influyente que no permitió que disminuyera el fraude por la modalidad FBM, como ya se había mencionado anteriormente; en la actualidad de acuerdo con las normas emitidas por las franquicias para Octubre del año 2016 los comercios debían contar con dicha tecnología en caso contrario el comercio asume el riesgo y se hace responsable del fraude, lo cual permitiría inferir que esta modalidad para el año 2016 y en adelante no presente mayor relevancia.

Capítulo 3. Marco Teórico

El interés en el estudio y análisis del fraude con tarjetas se ha incrementado en los últimos años debido a que son los medios de pago más utilizados en el país y por ende requiere de un estudio adecuado, día a día profesionales de diversas áreas han enfocado este estudio desde diferentes metodologías y puntos de vista.

En este sentido, este documento propone utilizar como técnica estadística, el análisis de series de tiempo, “La identificación y estimación de modelos de series temporales han sido desarrollados para procesos estacionarios” (Fernández, en línea, 2011, p.10). Estos modelos se clasifican en Autoregresivos AR (p), medias móviles MA (q), o el producto de su combinación autoregresivos de medias móviles ARMA (p, q).

Dicha técnica permite la predicción del comportamiento de una variable a futuro, una serie tiempo es una secuencia de datos u observaciones recopilados en intervalos de tiempo, para este caso la serie es mensual a partir de enero del año 2011 a diciembre del año 2015, organizados cronológicamente y los valores que toma la variable en observación tiene en cuenta tres componentes; componente tendencia, el cual se puede definir como un cambio a largo plazo y consiste en las variaciones de la serie McCleary y Hay definen la tendencia como:

Cualquier cambio sistemático en el nivel de una serie temporal, este componente puede entenderse como un movimiento determinístico, suave, amplio y no oscilatorio que se extiende en una determinada dirección, en sentido positivo, hacia arriba, o en sentido negativo, hacia abajo (Jara y Rosel, 2002, p. 22).

El componente estacional, variabilidad en los datos debido a influencias de las estaciones, efectos que hacen que la serie tenga en ciertos periodos de tiempo el mismo comportamiento “La estacionalidad se define por una fluctuación cíclica o periódica de la serie temporal que se repite de forma regular” (Arnau, 2001, p. 92) y el componente aleatorio, factores a corto plazo impredecibles o fortuitos que alteran la serie.

Como se mencionó anteriormente el principal objetivo de una serie de tiempo X_t donde $t = 1, 2, 3, 4$. Es su análisis para hacer pronóstico, con la finalidad de servir como guía para las decisiones que se deben tomar, en el caso que le compete a este documento el comportamiento del fraude con tarjetas para la entidad, las decisiones de prevenir y reaccionar ante el fraude con base en los pronósticos del mismo, de acuerdo a la participación de cada una de sus modalidades, es posible identificar una disminución o incremento en la tendencia del fraude lo que conlleva a implementar por la entidad políticas y técnicas con el fin de reducir y atacar con mayor oportunidad el riesgo.

Dado lo anterior para diseñar un modelo que permita predecir el comportamiento del fraude para la entidad financiera, cobra gran importancia realizar un análisis integral y grupal al comportamiento de estas variables y la magnitud e incidencia de las mismas. Por tal motivo se procede a realizar la estimación a través de un modelo econométrico que permita generar un análisis crítico, veraz y robusto, por lo anterior, se decidió utilizar las siguientes técnicas estadísticas divididas en tres fases, estimación del modelo ARIMA para la serie fraude total, utilización del método para realizar ajustes a los valores atípicos relevantes en la serie fraude total y por último se estima el modelo ARIMAX y se pronostica el fraude con tarjetas para 12 meses durante el año 2016.

De esta manera se realiza la estimación del modelo ARIMA su importancia radica en poder explicar el comportamiento y la evolución de la variable fraude total que coincide con un proceso no estacionario según los métodos de caracterización y evaluación que se aplican, simplemente porque cambia de nivel y de varianza con el tiempo por lo cual al proceso se le considera un proceso integrado, esto significa que la serie debe ser diferenciada d veces para lograr la estacionariedad, el resultado un modelo ARIMA (p, d, q) por sus siglas en inglés (Autorregresive integrated moving average) reúne la componente autorregresiva y de media móvil de la serie de tiempo, o en su defecto con el método **Box-Cox**.

Esto con el fin de obtener una serie de tiempo Y_t , es un conjunto de datos obtenidos en intervalos e tiempo con la misma periodicidad, donde Y_t se distribuye normalmente con media y varianza constante, $Y_t \sim (0, \sigma^2)$, para todo t y la covarianza entre Y_t y Y_s con $t < s$, depende del número de etapas que las separan, invariante en el tiempo.

La componente autorregresiva se encarga de determinar cada observación como una combinación lineal de las observaciones anteriores, mientras que la componente de media móvil incluye la parte aleatoria, P es el número de términos autorregresivos, d el número de veces que la serie debe ser diferenciada y q el número de términos de la media móvil.

Expresado en forma del polinomio operador de retardos el modelo ARIMA (p, d, q) es:

$$\Phi(L)(1-L)^d X_t = c + \theta(L)\varepsilon_t$$

Ecuación 1. Modelo ARIMA Expresado en forma del polinomio operador de retardos.

Donde X_t es la serie de diferencias de orden d , ε_t es un proceso de ruido blanco y $C, \Phi_1, \dots, \Phi_p, \theta_1, \dots, \theta_p$, son los parámetros del modelo.

Su construcción se hace mediante el siguiente proceso:

1. **Identificación:** Con base en los datos se propone el tipo de modelo ARIMA (p, d, q) esperando que p, d, q sean los apropiados para la estimación, aquí es posible identificar más de un modelo.
2. **Estimación:** una vez escogido el modelo se hace inferencia sobre sus parámetros.

3. **Validación:** se realizan pruebas o contrastes para verificar que el modelo se ajuste a los datos de la serie, de no ser así, se escoge un nuevo modelo y se repiten los pasos anteriores.
 4. **Predicción:** Una vez seleccionado el mejor modelo se pueden hacer pronósticos.
- Esta clase de modelos representa claramente las componentes de tendencia (U_t) y de estacionalidad (r_t) tal que la variable dependiente se obtiene como la suma de las dos anteriores.

$$Y_t = u + r_t + \varepsilon_t,$$

Donde ε_t Representa los errores del modelo y contienen relaciones que no alcanza a capturar el modelo.

También es importante en la metodología para la construcción de un modelo ARIMA, La evaluación previa a partir de gráficos con el propósito de identificar estacionalidad o estacionariedad en la serie de datos, por medio de las gráficas de autocorrelación simple (ACF) y parcial (PACF), hacer la transformación correspondiente si la serie de datos no es estacionaria: d si la serie es no estacionaria, D si no lo es con respecto a las correlaciones, Box Cox si no lo es con respecto a la varianza.

La estimación se realiza mediante algoritmos de máxima verosimilitud, si los parámetros son estadísticamente significativos, se especifican los residuales del modelo, si no hay tendencia en la función de autocorrelación de los residuales y las estadísticas de Box-Ljung en cualquier rezago no son significativas, el modelo es adecuado y por último se escoge el modelo final con base en un criterio de información, que mida en conjunto el error de ajuste de los datos y el error de varianza inducido por los grados de libertad del modelo, el criterio de información Akaike (AIC) será útil en esta tarea y se escoge el de menor AIC o menor BIC.

Uno de los posibles problemas es determinar el orden de integración de los datos, ya que es difícil distinguir entre un proceso autorregresivo y un proceso que sigue un paseo aleatorio, para la solución del problema se plantea una prueba estadística (Dickey Fuller) que determine cuando el coeficiente del modelo autorregresivo (p) es significativamente diferente de la unidad, para lo cual se realiza una regresión (Hernandez, Velasquez y Dyner, 2005, P. 22):

$Y_t = u + \beta * t + p * y_{t-1} + e_t$ “Siendo u el término de media constante, β la pendiente, p es el parámetro autorregresivo y e ruido blanco, y_t sigue un proceso integrado si $p=1$. La prueba asume $H_0: p=1$, contra $H_a: p < 1$ ” (Dickey Fuller, 1979).

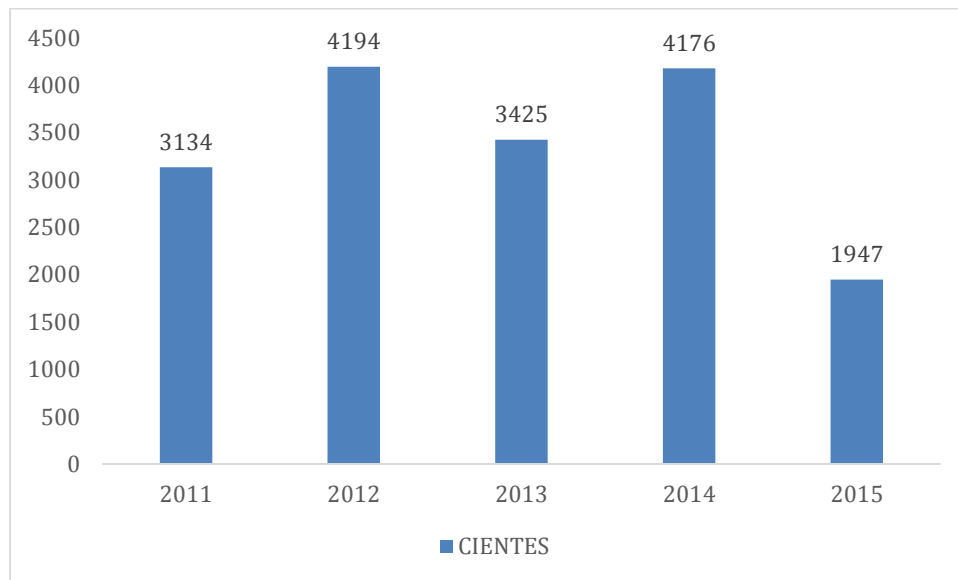
Para el modelo ARIMAX analizaremos 7 series temporales, las series aludidas son la recopilación del fraude con tarjetas para la entidad, por ser un modelo multivariado recibe el nombre de ARIMAX, con el fin de diferenciarlo del modelo univariado ARIMA, y podría indicarse de la siguiente manera $Y_t = f(X_1, X_2, X_n) + \varepsilon_t$, donde Y_t es la variable dependiente, X_n son las variables independientes y ε_t es el componente de error, este modelo tiene la ventaja de considerar varias variables.

Capítulo 4. Marco Metodológico

Antes de dar inicio al desarrollo del modelo es importante definir la metodología a utilizarse que permita cumplir con los propósitos de este documento. Frecuentemente se encuentran análisis econométricos teniendo como base la metodología de series de tiempo la cual permite ver el comportamiento de un individuo en relación a unas variables asociadas y su comportamiento a través del tiempo; es por ello que se propone realizar un análisis a través de modelo ARIMA y ARIMAX los cuales analizan varios individuos (en este caso las modalidades de fraude de la entidad financiera) en diferentes periodos de tiempo y asociados a la variable de interés fraude total.

Unidad de Análisis

Para poder llevar a cabo estas estimaciones se tomo como fuente de recolección de información, los datos mensuales de fraude que tiene la entidad en estudio por cada una de sus modalidades antes mencionadas (FMB, IN, S, NO F, SU-RO-EXTR y USO INDEBIDO), esto debido a que la ocurrencia de un fraude por una modalidad no implica el fraude por otra modalidad, las normas y sanciones expuestas anteriormente y con el cambio de tecnología migrada a Chip, se decide tomar la serie durante el periodo comprendido 2011- 2015 en el cual fueron afectados 16.876 tarjetahabientes de la entidad financiera por un valor \$19.033 Millones de pesos, discriminados de la siguiente manera:



Fuente: Entidad Financiera, Elaborada por el autor.

Figura 2: Clientes afectados por Fraude.

En la figura 2 se puede observar los clientes afectados por fraude durante el 2011 – 2015, siendo los años 2012 y 2014 los años con más tarjetahabientes afectados, en el 2015 se observa una disminución del 46% con respecto al año inmediatamente anterior, lo que permite identificar que la entidad está tomando mejores prácticas para evitar la continuidad del fraude y evitar que sus clientes sean afectados,

entre estas políticas se presentan la afinación de reglas que permitan generar alertas de fraude, investigación del 100% de los casos reclamados como fraude con el fin de evitar reportar a las franquicias MasterCard y Visa transacciones realizadas por los clientes, esto debido a que en años anteriores esto no se realizaba y a partir de la nueva metodología adoptada por la entidad se han identificado casos que aun cuando no corresponden a fraude se les han pagado a los tarjetahabientes y por ende reportado, esto en los últimos 5 años ha costado \$490 millones de pesos con una participación del 2% con respecto al fraude total; este modelo con base en lo anterior cobra especial relevancia para la entidad con el fin de predecir el posible fraude durante los periodos siguientes e identificar la concentración del fraude con el fin de reducirlo teniendo en cuenta la concentración del mismo en un 70%.

De otra manera es necesario identificar el enfoque de la investigación es cuantitativo porque El investigador o investigadora plantea un problema de estudio delimitado y concreto y sus preguntas de investigación tratan de cuestiones específicas, se interpreta a la luz de predicciones iniciales y esta constituye una explicación de cómo los resultados encajan con el conocimiento existente. Además, sigue rigurosamente procesos estadísticos y de acuerdo con ciertas reglas lógicas, los datos generados poseen los estándares de validez y confiabilidad, y las conclusiones derivadas contribuirán a la generación de conocimiento (Hernandez, Quinta Edición, 2005, p. 5).

Con base en lo anterior el alcance del proyecto es descriptivo,

Busca especificar las propiedades y características de un fenómeno que se somete a un análisis describiendo sus dimensiones y tendencias, es decir, únicamente pretenden medir o recoger información de manera independiente o conjunta sobre los conceptos o las variables a las que se refieren, el estudio tiene como propósito conocer la relación o el grado de asociación que existe entre variables en un contexto particular. La utilidad principal que ofrecen estos estudios es saber cómo se puede comportar una variable al conocer el comportamiento de otras variables vinculadas. Es decir, intentar predecir el valor aproximado que estas tendrán (Hernandez, Quinta Edición, 2005, p. 7).

Instrumentos

Ahora bien para llevar a cabo el modelo se utiliza como instrumento de investigación el uso del software estadístico R-Project, que es un software libre, de código abierto disponible en <http://www.rproject>, El software R a través del uso de librerías especializadas que respaldan su funcionalidad, es actualmente una herramienta valiosa en el desarrollo de modelos estadísticos que, dentro del contexto no solo académico, apoya la modelación de series económicas con variables que presentan usualmente cambios en el tiempo, este software presenta una amplia variedad de pruebas para la validación de supuestos y la capacidad de pronósticos.

Procedimiento

Inicialmente se realiza un análisis de raíces unitarias que permiten validar el comportamiento y estacionariedad de las series utilizadas para el desarrollo del modelo; posteriormente en busca de observar los datos atípicos que presenta la serie y poder validar a que corresponden así como ver la

participación de las modalidades de fraude con respecto al fraude total se utiliza la metodología desarrollada por Box y Jenkins en series de tiempo (1970), y culminando con la predicción del fraude para el año 2016; que más allá de constituir un ejercicio académico, pretende aportar una nueva herramienta de evaluación del fraude fiable, que permita un pronóstico en el comportamiento del fraude con sus modalidades y brinde elementos claves para el diseño de políticas enfocadas a prevenir y reaccionar en el riesgo con tarjetas de crédito que podría presentar la entidad.

En este documento se analizan las cifras correspondientes al fraude ocurrido de acuerdo con sus modalidades (FMB, IN, S, NO F, SU-RO-EXTR y USO INDEBIDO), con el fin de estimar el modelo ARIMAX en el periodo comprendido entre 2011 – 2015.

A través de este modelo se pretende estimar el coeficiente (β_i) en la siguiente ecuación:

$$(0) \quad FT_t = \beta FBM_t + \beta IN_t + \beta S_t + \beta NO F_t + \beta SU-RO-EXTR_t + \beta USO INDEBIDO + \varepsilon_{it}$$

Ecuación 2. Estimación Coeficiente (β_i)

Aquí se emplean el subíndice $t = 1, 2, 3, 4, \dots, T$. para denotar el tiempo; además se identifica FT_t como el fraude total de la entidad y los demás componentes teniendo en cuenta cada modalidad de fraude.

Cabe resaltar que para el desarrollo del modelo ARIMAX se debe tener en cuenta que el término de error está compuesto por tres partes:

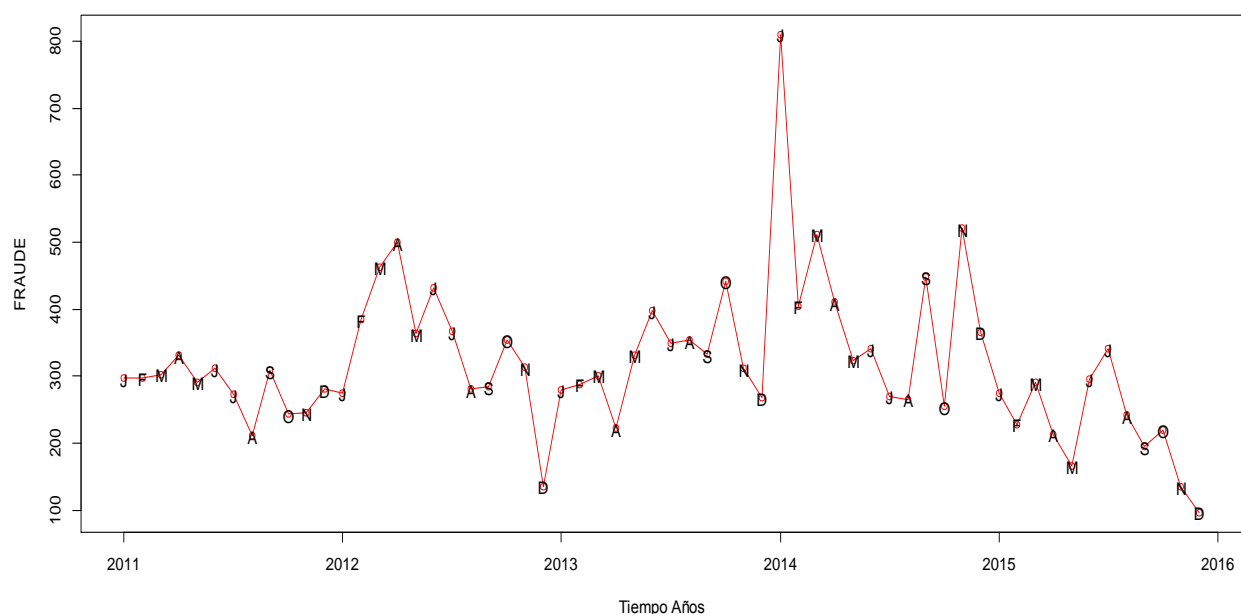
$$U_{it} = \alpha_i + \theta_i + Eu$$

Ecuación 3. Terminio de Error

Donde α_i es el componente individual que no varía a través del tiempo, θ_i es el factor temporal que no varía entre los individuos y Eu representa el efecto de todas las variables que componen el modelo y que cambian entre los individuos y el tiempo.

Capítulo 4. Análisis y Resultados

La serie de tiempo utilizada corresponde al fraude con tarjetas reportado por la entidad ante las franquicias MasterCard y Visa según sus modalidades, esta serie esta expresada en millones de pesos, la serie de tiempo contiene 60 datos que representan la información mensual comprendida entre el año 2011 y 2015, la cual fue obtenida de las bases de datos de reporte de la organización.



Fuente: Entidad Financiera, cifras de fraudes expresados en Millones.

Figura 3: Comportamiento del Fraude mensual entre el periodo 2011 – 2015.

La figura 3 muestra la evolución del fraude para la Entidad financiera durante el periodo comprendido entre el 2011 y 2015, un análisis grafico podría indicar que durante el periodo 2011 la serie muestra un comportamiento homogéneo, sin embargo; para los siguientes años se llega al reconocimiento de varios datos atípicos, el primero en Diciembre del 2013, el segundo en Enero del año 2014, con un considerable aumento representado en 810 millones de pesos, este dato atípico obedece al resultado de un punto de compromiso de información que afecto considerablemente a la Industria y como tal a la entidad financiera en cuestión, siendo materializado bajo la modalidad de fraude por Internet con una participación del 70%. También se alcanza a distinguir que entre septiembre y noviembre del mismo año, hay incrementos considerables en el fraude reportado, el cual también corresponde al punto de compromiso del mes de enero, sin embargo en esta segunda ocasión la Entidad Financiera se encontraba con una mayor capacidad de reacción, lo cual permitió que la dimensión del fraude no fuera igual a la ocurrida en enero del 2014, es de señalar que en el año 2015 la serie parece mostrar una disminución paulatina para todas las modalidades en comparación con los años anteriores, lo cual es el resultado de mejores prácticas de control por parte del banco.

No obstante, la serie es el resultado de nuevas tendencias en el uso y manejo del tarjetahabiente y el comportamiento del fraude, a continuación, se presentan un resumen y descripción de algunas medidas de estadística descriptiva asociadas con la serie:

Tabla No. 2
Análisis descriptivo

Descripción	Resultado
Min.	96.09
Cuartil 1	266.97
Median	299.32
Mean	317.23
Cuartil 3	356.53
Max.	810.03
Des. Estándar	109.4265

La tabla 2 presenta el resumen de estadística descriptiva de la serie fraude de la entidad financiera, teniendo en cuenta la Desviación Estándar la cual es de 109,43, se puede observar que esta es muy grande si se toma como punto de referencia la media. Esta situación haría que se presenten dificultades si se trata de predecir, solamente con el promedio, futuros valores de fraudes, ya que los datos presentan una gran dispersión. Lo anterior evidencia la necesidad de generar un modelo avanzado que permita predecir futuros montos de fraude con mayor exactitud que solamente usando la media.

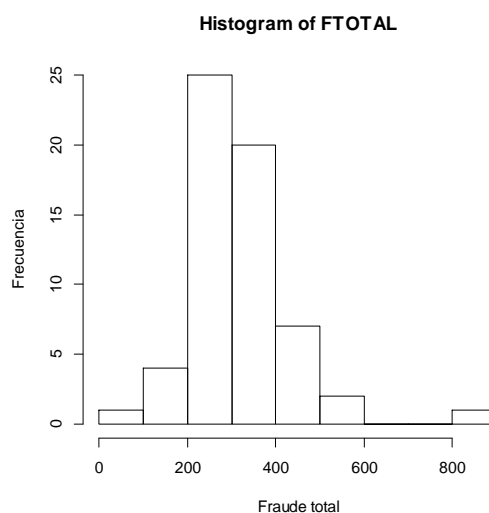


Figura 4: histograma con función de densidad de la serie de Fraude del Entidad financiera.

De acuerdo a la Figura 3 los datos presentan un sesgo hacia la derecha, lo que supondría que en los años del 2012 al 2014 se presentó una mayor concentración de los valores de la serie.

Inicialmente se toman las series en niveles para estimar la estacionariedad de las mismas y poder identificar el mejor modelo, se debe conocer si la serie es o no estacionara, para este fin se utiliza el test de Dickey – Fuller. El test arroja los siguientes resultados:

Dickey-Fuller = -2.1086, Lag order = 3, p-value = **0.5309**

alternative hypothesis: stationary

Considerando el P valor de 0.5309 no se rechaza la hipótesis nula, se parte de la modelación de la serie fraude a través de un esquema de series de tiempo autorregresivo, y por lo anterior es necesario determinar el orden de integración de la serie, debido a que esta no es estacionaria y se debe trabajar con diferencias, con el fin de evitar el problema de regresión espuria y no incurrir en el error de realiza un modelo con variables no estacionarias y por ende generar estimadores ineficientes y resultados erróneos.

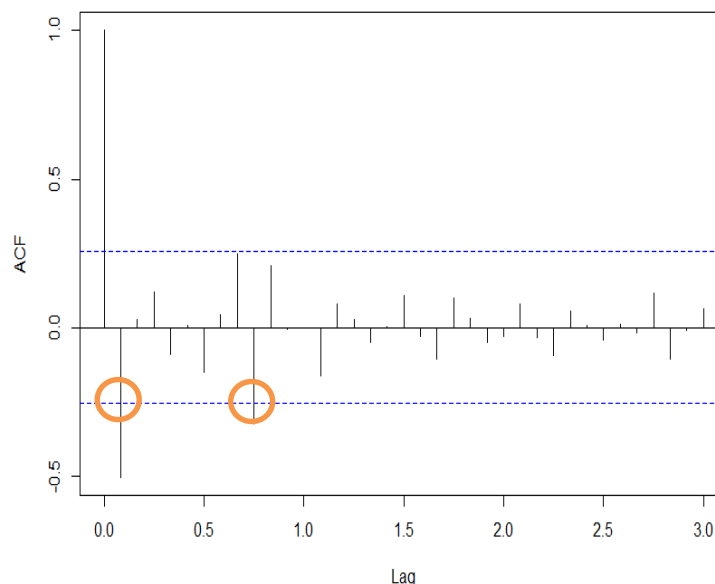
Por lo anterior se procede entonces a realizar el cálculo teniendo en cuenta como base la serie en su primera diferencia, el test arroja los siguientes resultados:

Dickey-Fuller = -4.4582, Lag order = 3, p-value = 0.01

alternative hypothesis: stationary

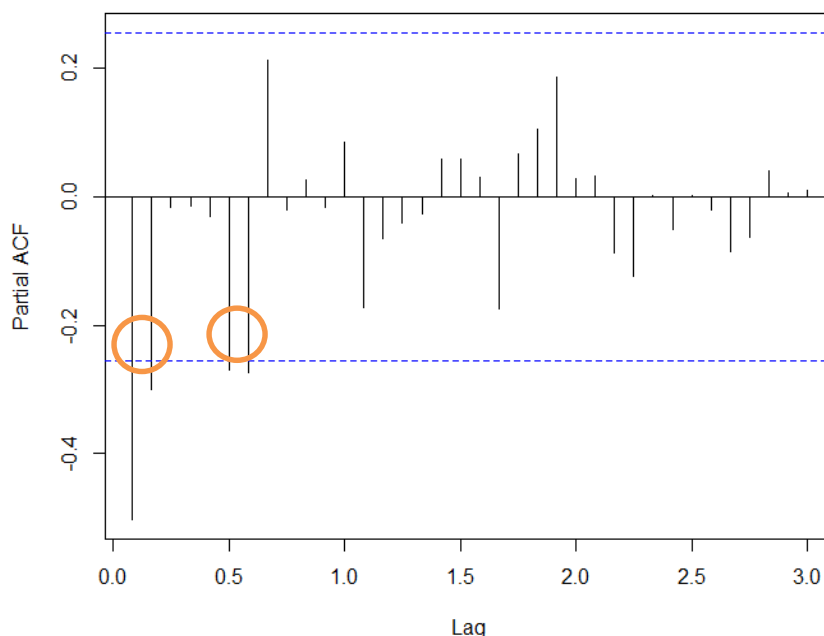
Se evidencia que cuando se toma la primera diferencia de la variable FT, se puede rechazar la hipótesis nula (H_0), la cual indica que la serie no tiene más raíces unitarias, esto debido a los datos obtenidos con los estadísticos, permitiendo establecer la estacionariedad de la variable.

Continuando con el análisis, al obtener el rechazo de la hipótesis nula, se comprueba que los residuos del modelo están exentos de más raíces unitarias, es decir, las pruebas de raíces unitarias permitieron establecer que la serie es integrada de orden 1, y es posible continuar con el modelo, para ello se realiza el grafico de la función de autocorrelación **ACF**.



Fuente: Elaboración propia a partir de los datos tratados en el programa estadístico R
Figura 5: Función de Autocorrelación

La figura 5 muestra la función de autocorrelación ACF muestra los Lags que exceden las bandas de significancia (1 y 9). De aquí se podría inferir que existe un modelo ARIMA (0, 1, 1), luego de validar la función de autocorrelación ACF es necesario validar la función de autocorrelación parcial PACF.



Fuente: Elaboración propia a partir de los datos tratados en el programa estadístico R

Figura 6: Función de Autocorrelación Parcial.

En la figura 6 se observa que los Lags que exceden las bandas de significancia: 1,2, 6 y 7. Es importante destacar que luego del segundo lag, los lags 3,4 y 5 se mueven dentro de las bandas de significancia.

Debido a que el modelo se trabaja con diferencias, se trabajara con un modelo ARIMA del tipo (p,1,q). Observando el comportamiento de los lags en la función de autocorrelación y autocorrelación parcial. Se proponen 4 posibles modelos idóneos:

1. ARIMA (2,1,0)
2. ARIMA (0,1,1)
3. ARIMA (9,1,7)
4. ARIMA (5,1,1)

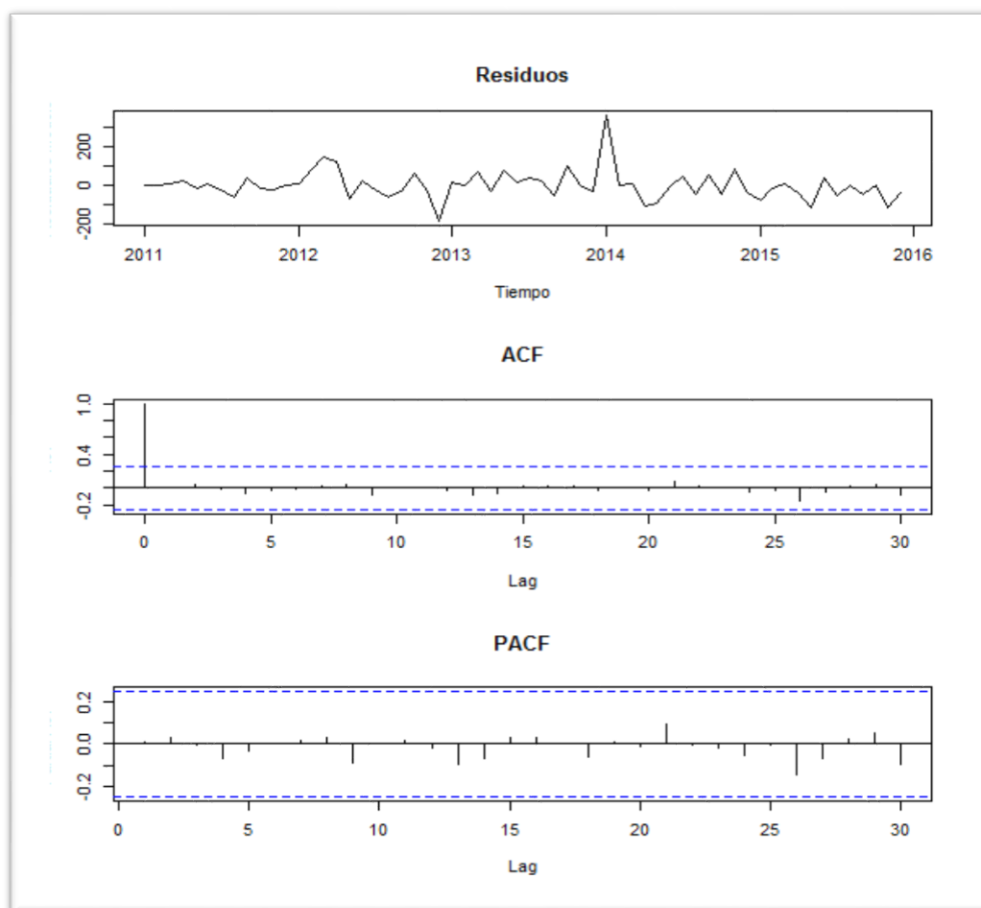
Analizando estos modelos, obtenemos los siguientes resultados estadísticos y pruebas para determinar cuál podría ser el modelo ARIMA a estudiar, para la serie fraude total con tarjetas de la entidad en estudio.

Tabla 3.

Descripción de los posibles modelos.

Modelo	AIC	BIC	Residuales (p-valor)		
			Box-Ljung	Jarque Bera	Shapiro-Wilk
(2,1,0)	720.46	726.69	0.4427	2.2e-16	1,80E-02
(0,1,1)	718.58	722.73	0.6521	2.2e-16	6,83E-03
(9,1,7)	728.22	763.54	1	2.2e-16	1,78E-02
(5,1,1)	723,34	739.882	0.7507	2.2e-16	3.94 e-06

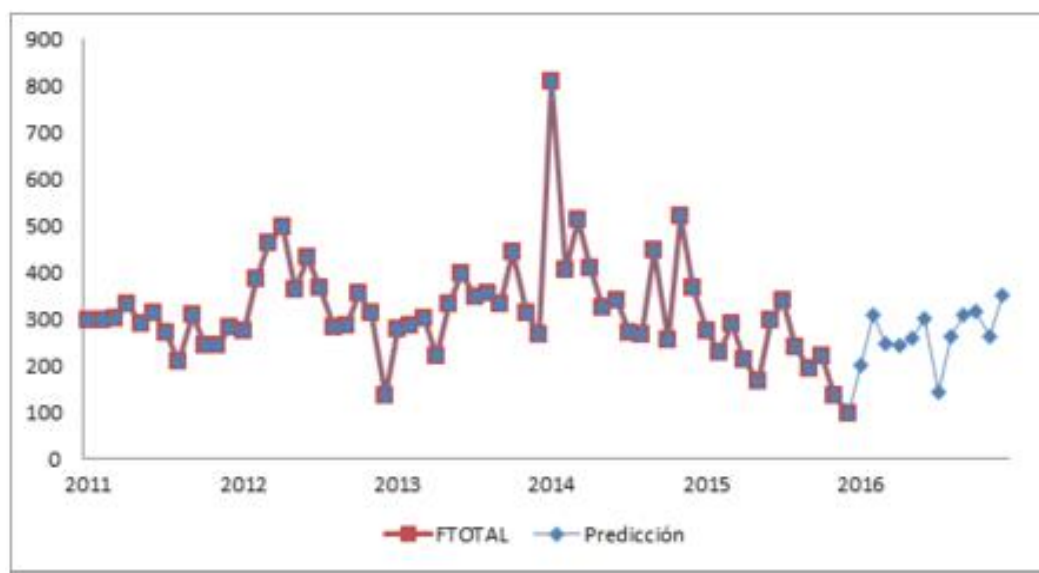
En la Tabla 3 observamos que el mejor modelo tanto por AIC como por BIC es el modelo ARIMA (0,1,1). Los residuales en este modelo se comportan con normalidad como permite observar los 3 test.



Fuente: Elaboración propia a partir de los datos tratados en el programa estadístico R

Figura 7: Análisis de residuales del modelo.

En la figura 7 se observan los residuos, y las gráficas de la función de autocorrelación y autocorrelación parcial, se puede concluir que el modelo no presenta ninguna clase de problema con los residuales, es posible notar un modelo sólido.



Fuente: Elaboración propia a partir de los datos tratados en el programa estadístico R

Figura 8: Ajustado Observado y Predicción.

En la figura 8 es posible ver la serie ajustada en comparación con la serie observada de fraude, es importante notar el ajuste que se logra con el modelo planteado, incluso en datos que presentan gran volatilidad.

Tabla 4.

Predicción del fraude para el 2016 Modelo ARIMA

Mes	Montto(Mill)
Enero	199.51
Febrero	307.17
Marzo	246.12
Abril	244.10
Mayo	260.83
Junio	300.67
Julio	144.95
Agosto	261.55
Septiembre	309.31
Octubre	316.15
Noviembre	261.12
Diciembre	351.07

La tabla muestra los valores mensuales de la predicción del fraude total para la Entidad financiera en el año 2016, los meses de septiembre a Diciembre muestran valores considerables, lo que es normal de acuerdo a las fechas de más facturación debido a la celebración del amor y la amistad en septiembre y la proximidad de la navidad y año nuevo; sin embargo es necesario en el análisis de series de tiempo realizar el suavizado de la serie, con base en los datos atípicos identificados denominados Outliers; los outliers son observaciones que se desvían de la serie de datos aumentando la varianza y usualmente afectando la precisión de las pruebas estadísticas como las de normalidad. Como se utiliza un modelo ARIMA y teniendo en cuenta que estos se usan para modelar datos con cierta homogeneidad, los outliers influyen en la bondad de ajuste del modelo de Fraude total.

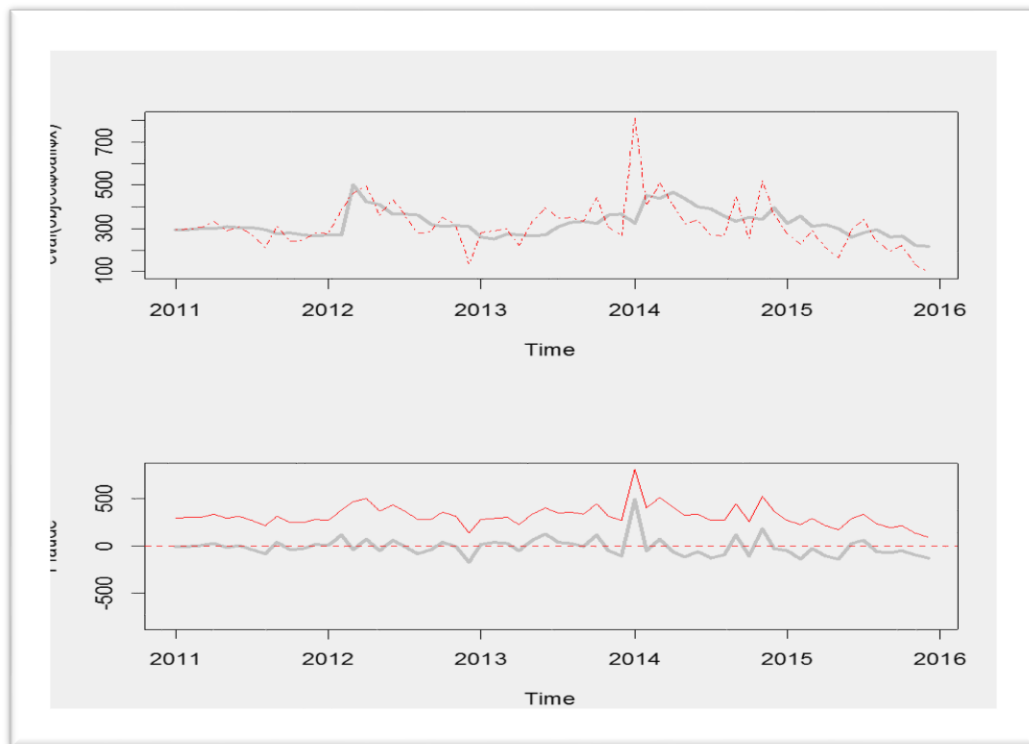
Tabla 5.

Ajuste Datos Atipicos Outliers (“AO”, “LS”, “TC”)

type	ind	coefhat	tstat
AO	37	4.456.724	6.193.853
AO	47	2.145.957	3.066.971
TC	15	1.926.244	3.093.797

En la tabla 5 se puede observar que se identificó la presencia de dos tipos de outlier aditivo (AO) y cambio temporal (TC), la identificación de estas observaciones significativamente diferentes se hace bajo el análisis de intervención y datos atípicos bajo modelos Box – Jenkins.

En la primera etapa de identificación de los outliers, consiste en ver en el modelo ARIMA los potenciales datos atípicos, el método de análisis de la serie temporal necesita que la serie sea estacionaria, por lo cual, se debe diferenciar procurando esta condición y se verifica a través de pruebas como la de Dickey-Fuller para la serie fraude total, esta es integrada de orden 1, una vez identificadas estas observaciones se realizan pruebas de significancia, de tal modo que el modelo recoja el ajuste, el software R identifico 3 datos atípicos identificados en los meses 37,47 y 15, sin embargo finalmente dejo los meses 47 y 15 y los cuales se tendrán en cuenta para el modelo ARMAX, teniendo en cuenta la serie fraude total ajustada en estos meses.



Fuente: Elaboración propia a partir de los datos tratados en el programa estadístico R

Figura Serie 9: inicial y con datos ajustados.

En la figura 9 es posible ver los datos de la serie ajustada en comparación con la serie observada de fraude, aquí se observa una gráfica que permite suavizar los datos atípicos encontrados en la serie Fraude total y que pueden estar generando problemas en el momento de la predicción y ejecución del modelo ARIMAX.

Tabla 6.

Estimación Modelo ARIMAX

`arimax(x = ftotal, order = c(0, 1, 1), xreg = data.frame(nregs1))`

Coefficients:					
ma1	fbm	In	nof	s	suroextr
-0.9421	10.282	0.9925	11.225	10.091	0.9840
s.e. 0.1509	0.0113	0.0065	0.1036	0.0434	0.0327

sigma² estimated as 14.99: log likelihood = -164.68, aic = 341.35

En la tabla 6 se puede observar que al incluir a la serie fraude total las modalidades de fraude, estas son altamente significativas para el modelo a excepción de la modalidad uso indebido la cual se retiró del modelo, debido a que no era altamente influyente en el fraude total, al incluirle variables independientes al modelos, se observa también que el criterio AIC, es más eficiente y por ende este

modelo reemplazaría al modelo ARIMA (0,1,1), por otra parte se comprueba que las modalidades son independientes y que la ocurrencia de un fraude bajo alguna modalidad no implica que la entidad tenga fraude con otra modalidad.

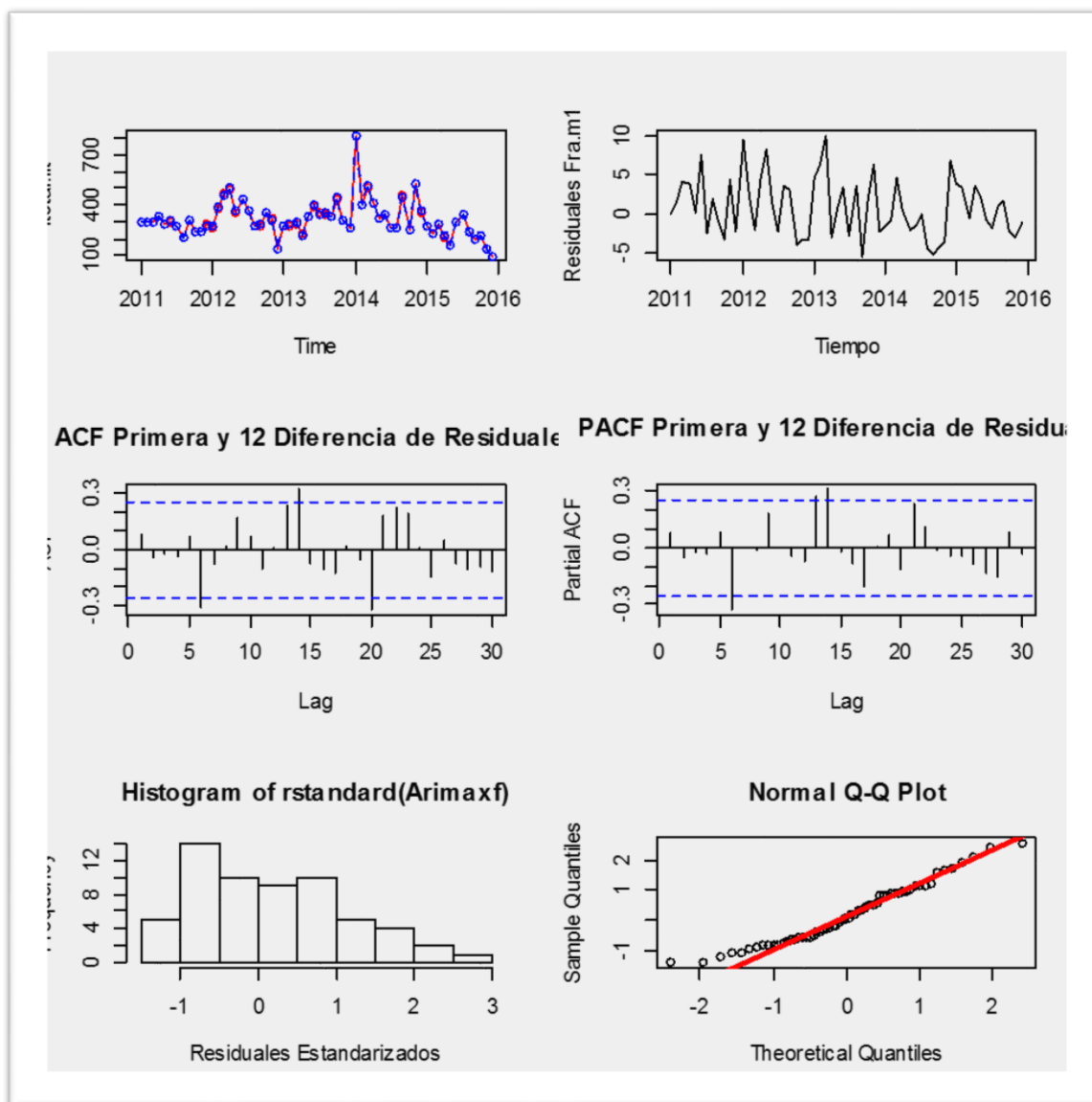


Figura 10: Resultado grafico del Modelo ARIMAX

La figura 10 muestra el resultado visual del modelo ARIMAX, con el ajuste de datos atípicos.

Tabla 7. Pruebas realizadas modelo ARIMAX

Box-Ljung test
data: as.vector(residuals(Arimaxf)) X-squared = 47.701, df = 24, p-value = 0.3145
Jarque Bera Test
data: as.vector(residuals(Arimaxf)) X-squared = 2.9747, df = 2, p-value = 0.226
Shapiro-Wilk normality test
data: as.vector(residuals(Arimaxf)) W = 0.9613, p-value = 0.05448

En la tabla 7 se puede observar las pruebas Box-Ljung Ho: Los datos no están correlacionados Ha: Los datos están correlacionados **P valor = 0.3145**, en este caso se aprueba Ho. La prueba Jarque Bera el **P valor = 0.226** y la prueba de Shapiro Wilk **P valor = 0.054**. Ho: Datos distribuidos normalmente y Ha: Los datos no están distribuidos normalmente, en este caso se aprueba Ho.

Tabla 8.

Proyección del fraude para el año 2016 en la entidad financiera, 32.

Predicción	
Jan	212.0712
Feb	227.9162
Mar	249.0872
Apr	315.8231
May	257.2276
Jun	325.4553
Jul	324.2937
Aug	272.4077
Sep	250.5957
Oct	285.1820
Nov	244.7790
Dec	236.9578

En la tabla 8 se observa la proyección del fraude para la entidad, siendo Junio el mes que presenta mayor fraude la entidad, y teniendo como proyección de fraude a cierre diciembre 31 del 2016 por un valor de 3.201 millones de pesos.

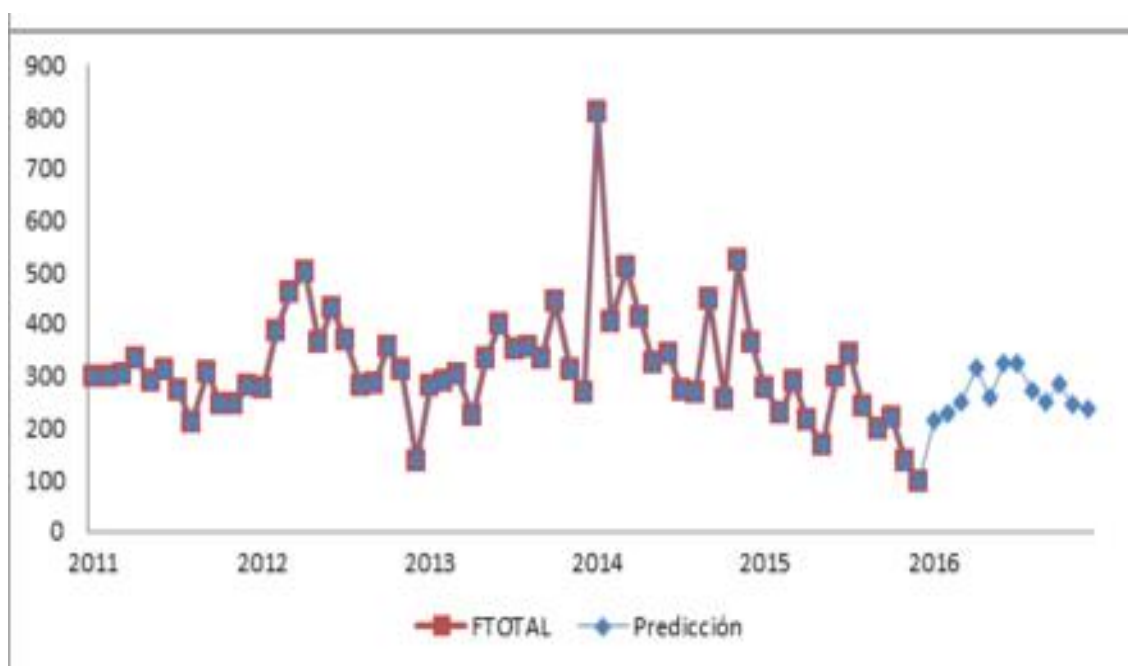


Figura 11: Resultado grafico de la Predicción del fraude para el 2016

La figura 11 muestra la predicción del fraude para el año 2016, siendo abril, junio y julio los meses en que se presenta mayor fraude.

Tabla 9.

Proyección de fraude v/s cifras de la entidad.

Mes	Predicción	Cifras Entidad
Jan	212,0712	371,759
Feb	227,9162	258,681
Mar	249,0872	116,807
Apr	315,8231	
May	257,2276	
Jun	325,4553	
Jul	324,2937	
Aug	272,4077	
Sep	250,5957	
Oct	285,182	
Nov	244,779	
Dec	236,9578	

En la tabla 9 se observa la proyección del fraude para la entidad y la comparación del fraude de los 3 primeros meses reportados por la entidad a las franquicias MasterCard y Visa, esto debido a que los clientes aún siguen reclamando transacciones y los datos de los meses Abril y Mayo están en proceso

de recolección, en el mes de Enero se muestra una diferencia significativa entre el dato reportado a las franquicias versus el pronóstico por valor de \$100 millones de pesos, en el mes de febrero se observa una diferencia tan solo de 30 millones de, este mes fue bastante acertado en el pronóstico realizado por el modelo ARIMAX, para el mes de marzo se vuelve a evidenciar una diferencia de aproximadamente \$100 millones con respecto al pronóstico y lo reportado por la entidad, sin embargo aquí es importante aclarar que la entidad para el mes de marzo no reporto ningún caso que correspondiera a no fraude, esto debido a que está en un proceso de atención de reclamos más riguroso y en su análisis está identificando oportunamente las transacciones realizadas por los tarjetahabientes y reclamadas como fraude y por ende no está permitiendo el reporte de las mismas a las franquicias, este tipo de procesos influyen significativamente en que el pronóstico realizado con el modelo ARIMAX no sea el más acertado con referencia al fraude reportado por la entidad, es también importante mencionar que la entidad en los años 2015 y 2016 ha implementado políticas importantes con respecto a prevenir el fraude y en la detección oportuna del mismo, aun cuando es también importante mencionar que para el mes de mayo aún sigue recibiendo reclamos de transacciones no reconocidas por los tarjetahabientes de los meses enero a abril.

Con base en lo anterior se decide validar la serie original ver si en los meses donde se presenta semana santa, junio y diciembre la serie presenta estacionalidad constante a través de los años, encontrando lo siguiente:

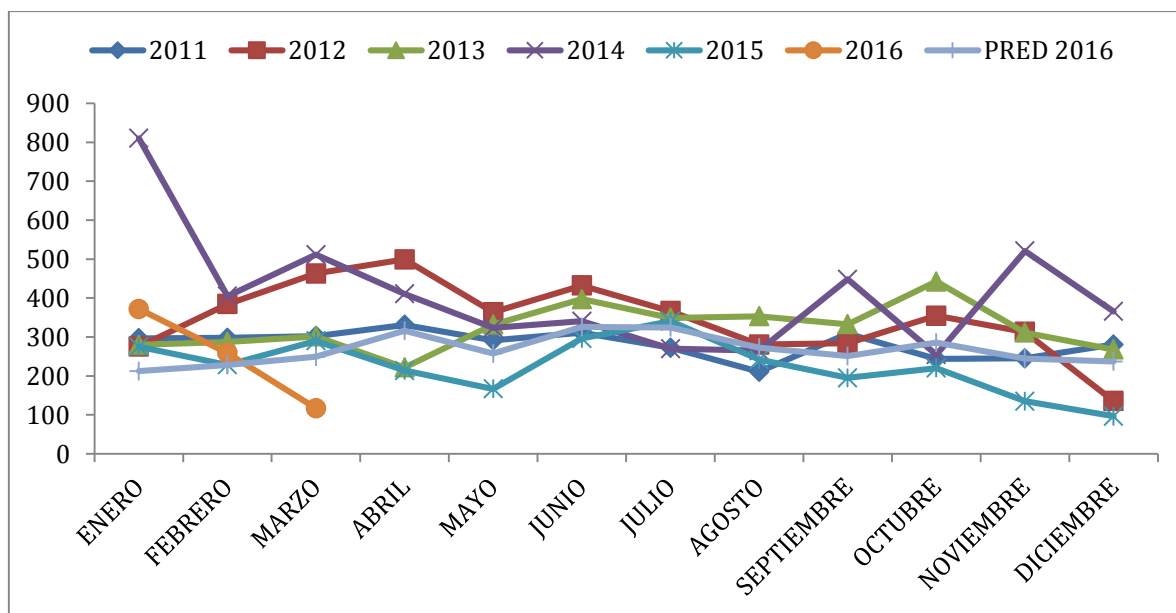


Figura 12: Predicción del fraude versus comportamiento mensual de la serie 2011-2016

La figura 12 se observa que la predicción otorgada por el modelo ARIMAX se encuentra en los rangos de confianza del comportamiento mensual de la serie, estando en un rango mensual entre 200 y

300 millones de fraude por mes, con base en lo anterior es necesario ver el promedio mensual de fraude que ha tenido históricamente la serie y si este se ajusta al modelo propuesto en este documento.

Tabla 10.

Promedio de fraude por mes y su proyección

AÑO	2011	2012	2013	2014	2015	2016	TOTAL	PROMEDIO	PRED 2016
ENERO	296,39	274,87	279,51	810,03	274,50	371,76	2307,05	384,51	212,07
FEBRERO	297,47	384,45	287,52	404,72	227,97	258,68	1860,81	310,14	227,92
MARZO	302,11	462,87	301,18	511,40	289,88	116,81	1984,25	330,71	249,09
ABRIL	330,92	499,20	221,77	410,56	213,63		1676,08	335,22	315,82
MAYO	290,88	363,28	331,88	323,17	166,48		1475,70	295,14	257,23
JUNIO	311,04	432,09	397,15	340,23	296,06		1776,57	355,31	325,46
JULIO	272,04	366,25	349,13	269,68	340,51		1597,61	319,52	324,29
AGOSTO	210,64	280,52	353,27	264,87	241,61		1350,91	270,18	272,41
SEPTIEMBRE	308,23	284,04	332,68	448,23	194,87		1568,05	313,61	250,60
OCTUBRE	243,63	354,28	442,26	254,36	219,90		1514,42	302,88	285,18
NOVIEMBRE	244,97	312,96	311,05	520,91	134,66		1524,56	304,91	244,78
DICIEMBRE	280,52	135,33	267,67	365,42	96,09		1145,03	229,01	236,96
TOTAL	\$ 3.388,87	\$ 4.150,14	\$ 3.875,06	\$ 4.923,57	\$ 2.696,17	\$ 747,25	\$ 19.781,04	\$ 3.751,14	\$ 3.201,80

En la tabla 10 se observa el comportamiento mensual de la serie año tras año y el promedio mensual de fraude durante los últimos 5 años versus el pronóstico del modelo ARIMAX (0,1,1), donde podemos observar que el modelo a partir del mes abril de 2016 es muy acertado versus el promedio de los otros meses 2011-2015, en este grafica también se observa que el promedio de fraude mensual esta entre 290 millones y 350 millones, no se observa incrementos fuertes en los meses de semana santa, junio o diciembre; el fraude es parejo para todos los meses.

Conclusiones

Fue necesario trabajar con las diferencias de la serie para conseguir estacionalidad ya que la serie original carecía de este componente, observando los gráficos de la función de autocorrelación parcial se plantearon tres posibles modelos ARIMA de los cuales se escoge el modelo (0,1,1) debido a que es el que presenta el AIC y BIC menor además, fue necesario probar normalidad sobre los residuos y finalmente se logra un buen ajuste entre la serie observada y la conseguida con el modelo.

Con base en el modelo ARIMAX la variable uso indebido no es significativa en la explicación del fraude total para la entidad.

Las variables que más contribuyen en la explicación del fraude total son fraude por Internet y fraude por Falsificación de banda magnética.

Con respecto a los valores que se observan en la predicción se estima que para el año 2016 se presentaran perdidas por un monto total de \$ 3201 millones, siendo junio el mes en que más dinero se perderá por fraude con un estimado de \$325 millones.

Con el análisis del modelo se espera que la entidad tome medidas orientadas a mitigar el fraude en pro de seguir manteniendo confianza corporativa esperando que esto converja en la obtención de una buena reputación frente a la industria y el cliente.

Como agenda de investigación consecuente con la línea de este documento se propone realizar un análisis de VAR que permita ver la predicción del fraude por cada una de sus modalidades con el fin de minimizar el impacto generado que tiene cada modalidad de fraude sobre el fraude total y ver el comportamiento y desarrollo de este proyecto cuya proyección de inicio está estipulada para inicios del presente año 2016.

También como agenda de investigación se puede estudiar y profundizar en el conocimiento de los factores que explican que un cliente tenga fraude dado que ha cumplido ciertas condiciones, es decir realizar un modelo segmentación del cliente y realizar un modelo basado en el éxito o fracaso, que sirva de ayuda para la entidad y permita preveer que clientes pueden llegar a tener fraude en el corto plazo.

Referencias Bibliográficas

Perote Peña J. (2011) Modelos de series temporales en finanzas (I) modelos ARIMA, Modelización económica II. Disponible en: Modelos ARIMA, consultado el 04 de Mayo de 2015 http://campus.usal.es/~ehe/perote/documentos/TEMA%202%20MODELIZACION%20ECONOMIC A%20II_2.pdf.

Hernandez N.O, Velázquez J. y Dyner I. (2005), Modelos ARIMA y estructura de la serie de precios promedio de los contratos en el mercado mayorista de energía eléctrica en Colombia. Universidad Nacional de Colombia, sede Medellín Colombia. Recuperado de: <http://www.bdigital.unal.edu.co/26445/1/24037-84126-1-PB.pdf>

Introducción a los Modelos ARIMA (2016), consultado el 04 de Mayo de 2016 <http://robjhyndman.com/hyndsight/arimaconstants/>

Banco de la republica (2016), Inclusión financiera en Colombia, en línea, citado el 12 de Mayo de 2016, disponible en Internet en: http://www.banrep.gov.co/sites/default/files/eventos/archivos/sem_357.pdf

Sistema financiero Colombiano, (2016) Superintendencia Financiera de Colombia, Dirección de Investigación y Desarrollo, Subdirección de análisis e Información.

MANNING, P.K. (2004), en Biblioteca digital, disponible en Internet en: <http://www.produccioncientifica.luz.edu.ve/index.php/capitulo/article/view/5220/5211>

Sistema financiero Colombiano, (2012) Superintendencia de Industria y Comercio disponible en internet en: <http://www.sic.gov.co/drupal/masive/datos/estudios%20economicos/Documentos%2020elaborados%20por%20la%20Delegatura%20de%20Protecci%C3%B3n%20de%20la%20Competencia/2012/5.%20Cr%C3%A9dito%20de%20Consumo%20DPC.pdf>

Sistema financiero Colombiano (2016), Superintendencia Financiera de Colombia, Dirección de Investigación y Desarrollo, Subdirección de análisis e Información. <https://www.superfinanciera.gov.co/jsp/loader.jsf?lServicio=Publicaciones&lTipo=publicaciones&lFuncion=loadContenidoPublicacion&id=10082699>

Estudio Universidad Nueva Granada (2014), las tarjetas de Crédito en Colombia, en línea, disponible en: <http://repository.unimilitar.edu.co/bitstream/10654/11837/1/Las%20Tarjetas%20de%20Credito%20en%20Colombia%20-%20Pedro%20Pablo%20Sanchez%20G.%20-%20Abr2014.pdf>

Abc de la circular externa 052 (2007), Superintendencia Financiera de Colombia, en línea, disponible en:

https://www.google.com.co/url?url=https://www.superfinanciera.gov.co/SFCant/ComunicadosyPublicaciones/ComunicadosdePrensa/abcce052_2007.doc&rct=j&frm=1&q=&esrc=s&sa=U&ved=0ahUKEwifra6F5OnMAhVEbR4KHV6mCQ4QFggZMAE&usg=AFQjCNFZzxfsigteQDCmy1xduQoX4NPk

Las tarjetas de Crédito en Colombia (2012), Reporte de modalidad según la norma emitida por Mastercard Diciembre del 2012.

Hernandez N.O, Velázquez J. y Dyner I. (2005), Modelos ARIMA y estructura de la serie de precios promedio de los contratos en el mercado mayorista de energía eléctrica en Colombia. Universidad Nacional de Colombia, sede Medellín Colombia. Recuperado de: <http://www.bdigital.unal.edu.co/26445/1/24037-84126-1-PB.pdf>

Hernández Sampieri, R. Definiciones de los enfoques cuantitativo y cualitativo, sus similitudes y diferencias. Quinta edición. Metodología de la investigación (PP 4-10), Bogotá, Colombia: Mc Graw Hill.

R-PROJECT: ENTORNO Y LENGUAJE (2016) Avances en Estadística Computacional, citado el 23 de Mayo de 2016, disponible en Internet en: <http://www.fce.unal.edu.co/uifce/proyectos-de-estudio/pdf/Rproject>

Sistema financiero Colombiano Superintendencia de Industria y Comercio (2012), estudios de mercado disponible en internet en: <http://www.sic.gov.co/drupal/masive/datos/estudios%20economicos/Documentos%20%20elaborados%20por%20la%20Delegatura%20de%20Protecci%C3%B3n%20de%20la%20Competencia/2012/5.%20Cr%C3%A9dito%20de%20Consumo%20DPC.pdf>